



Pergamon

PII: S0042-6989(96)00272-6

Vision Res., Vol. 37, No. 10, pp. 1281–1293, 1997
 © 1997 Elsevier Science Ltd. All rights reserved
 Printed in Great Britain
 0042-6989/97 \$17.00 + 0.00

Spatio-temporal Boundary Formation: the Role of Local Motion Signals in Boundary Perception

THOMAS F. SHIPLEY,*[‡] PHILIP J. KELLMAN[†]

Received 26 April 1995; in revised form 20 May 1996; in final form 20 September 1996

Spatio-temporal boundary formation (SBF) refers to a perceptual process responsible for perception of moving, bounded surfaces from sequential changes in spatially separated local elements. Previous research has indicated that this process produces perception of global form, continuous boundaries and global motion from spatially and temporally sparse element changes. In the present paper, we sought to distinguish between two classes of models for SBF: *form-precedes-motion* and *motion-precedes-form* models. Experiment 1 tested the effects of the addition of spurious motion signals, a manipulation that should affect a motion-precedes-form computation but not a form-precedes-motion computation. Shape identification in a 10-alternative forced-choice procedure was disrupted by this manipulation, supporting the former class of models. A particular computational scheme, edge orientation from motion (EOFM) instantiating a motion-precedes-form model is described and tested in Experiment 2. The EOFM model should be disrupted when initiating element changes occur in a certain type of sequential order, relative to randomly arranged changes. Sequential changes markedly disrupted performance, supporting this EOFM approach. The results favor motion-precedes-form models of SBF and are consistent with the particular computational scheme proposed. © 1997 Elsevier Science Ltd. All rights reserved.

Form Shape Motion Spatio-temporal filling-in

INTRODUCTION

As we move through our environment its visual projection constantly changes, yet the world we perceive is stable. Stable percepts might seem to depend on optical information that does not change over time. Research on visual surface segmentation has naturally focused on static sources of information, such as the discontinuities in luminance, color, texture, and stereoscopic disparity generally found at surface edges. But surface boundaries are not always specified by discontinuities in static properties, as when similar objects are clustered together (e.g. stands of trees). Moreover, over time the positions and configurations of boundaries defined by static properties change. These changes arise from several sources. Some are nonrigid deformations of surfaces, and many are changes in position of surface discontinuities resulting from object or observer movement.

Accurate segmentation of visual scenes despite movement-related changes may rely heavily on dynamic information. Gibson (1966, 1979) argued that patterns of change over time can provide information about persisting properties of the spatial layout. Gibson *et al.* (1969) offered an analysis of how patterns of change

might provide information for stable properties of the world. The pattern of deletion and accretion—the disappearance and reappearance of texture elements—that occurs when one object occludes and reveals another, provides information both for the continued existence of the occluded surface, and the shape of the nearer surface (Gibson, 1968). A number of researchers (Andersen & Cortese, 1989; Stappers, 1989; Bruno & Bertamini, 1990; Bruno & Gerbino, 1991; Shipley & Kellman, 1993c, 1994) have demonstrated that observers can use occlusion related events like texture disappearance and reappearance to perceive the boundaries of a moving figure. Figure 1(a) illustrates dynamic occlusion in a sparse random dot kinematogram. In such dynamic occlusion displays observers typically report seeing a moving form with well defined edges. Understanding the visual mechanisms responsible for producing an edge in these displays may be an important step in understanding how segmentation of boundaries occurs in rich natural scenes.

Shipley and Kellman (1993c, 1994) suggested that the appearance of edges defined by accretion and deletion of texture was one case of a general dynamic unit formation process, which they termed *spatio-temporal boundary formation* (SBF). Their studies used displays consisting of discrete texture elements, for example, small circles scattered on a homogeneous background. The initiating events for SBF are abrupt changes in single elements,

*Weiss Hall, Temple University, Philadelphia, PA 19122, U.S.A.

[†]Franz Hall, University of California, 405 Hilgard Ave, Los Angeles, CA 90024, U.S.A.

[‡]To whom all correspondence should be addressed.

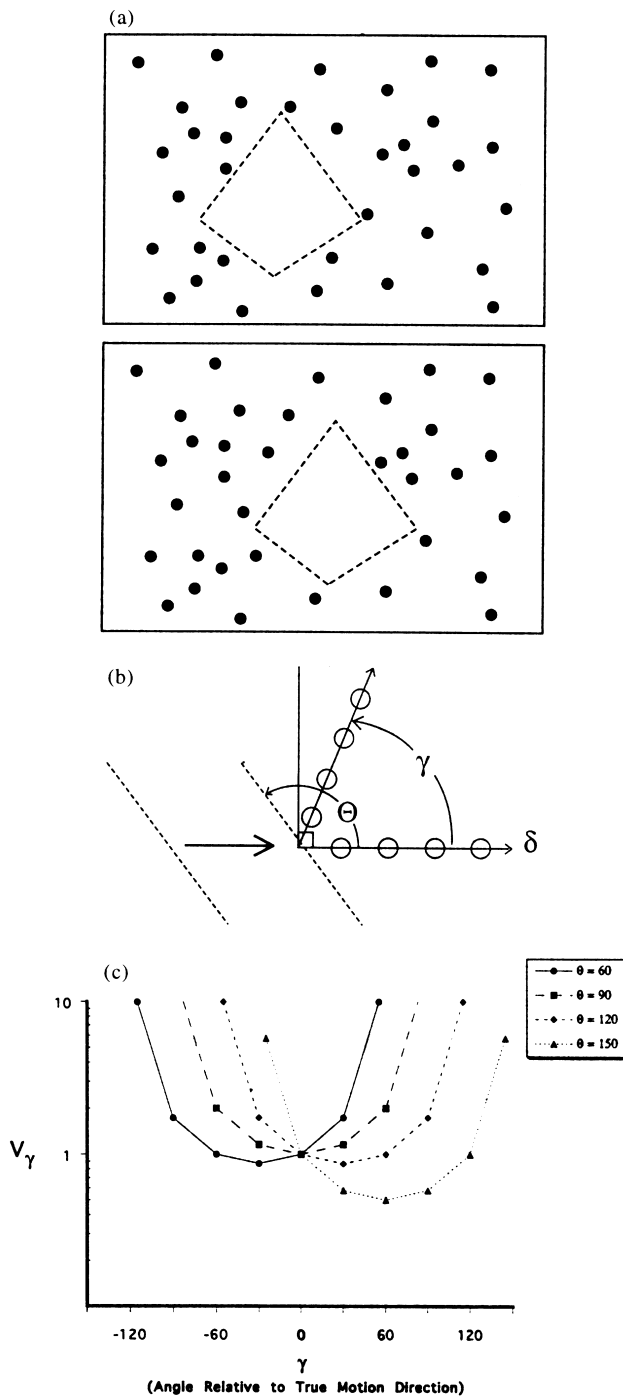


FIGURE 1. (a) An illustration of a white form (the dashed diamond) moving over a field of small black elements. In such dynamic occlusion displays elements at the leading edge disappear while elements at the trailing edge appear. (b) A diagram illustrating a pseudosurface edge moving in direction δ over one-dimensional strips of surface elements. The variable θ represents the orientation of the edge relative to the direction of motion, and the variable γ represents the orientation of each strip of elements relative to the direction of motion. (c) When edges with differing orientation (θ) move over one-dimensional element strips at various γ , the resulting velocities along each strip (V_γ) show a characteristic pattern. The minima for each curve occurs at $\theta = 90$ deg, and the intersection of two or more curves is the true direction of motion (d). [Fig. 1(b and c) are similar to Figs 12 and 13 from Shipley & Kellman (1994). Copyright © (1994) by the American Psychological Association. Adapted with permission.]

referred to as *spatio-temporal discontinuities* (STDs). In an SBF display, these changes are determined by the

position of a moving, invisible, mathematically defined form called the *pseudosurface*. Specifically, as the pseudosurface moves over the array, STDs occur for elements entering or leaving the pseudosurface.

Experiments showed that accretion and deletion of texture is part of a much larger class of local element transformations whose spatial and temporal arrangement give rise to perceived boundaries, form, and global motion (Shipley & Kellman, 1994). Such changes include transformations of color, orientation, form, and position. In dynamic occlusion displays [e.g. Figure 1(a)], the elements undergo *unidirectional transformations*. This term refers to the class of displays in which all elements have one value before entering the pseudosurface and switch to a different value upon entering. Upon exiting, the elements revert to their original value. *Bidirectional transformations*, in contrast, refer to cases in which elements in the array have either one of two initial values of some attribute; upon entering the pseudosurface, that attribute for each element changes to the other value. Again, each element reverts to its original value upon exiting the pseudosurface. An example of a bidirectional transformation would be a display containing blue and white dots on a gray background, in which entry into the pseudosurface region causes the blue dots to turn white and the white dots to turn blue. Such displays are both practically useful and theoretically important since no static property (e.g. luminance or texture difference) defines the forms seen in these displays (Shipley & Kellman, 1994).

In addition to a bounded form, motion is also perceived in SBF displays. Successive STDs, for example, color changes, in nearby locations would activate motion detectors (Chubb & Sperling, 1988; Shipley & Kellman, 1994); we refer to this triggering of motion-sensitive units as *local* motion signals, whereas we will refer to perceived motion of the pseudosurface as *global* motion.

Perceivers' abilities to detect object shape, continuous boundaries, and motion from sequential changes in spatially separate elements present a puzzle. Shipley and Kellman (1994) proposed two general classes of models to account for this ability. The two classes differ in the causal roles played by form and motion information. In a *form-precedes-motion* model, edges are constructed between element changes that occur closely in space and time. Subsequent element transformations may define this segment in a nearby position, allowing estimation of velocity. Overall form could be constructed by combining nearby edge tokens into larger units.

Alternatively, in a *motion-precedes-form* model, motion is derived in advance of boundaries and form. Recovering a form's global motion from individual motion signals requires solving an aperture problem. The solution might involve a constraint, such as the minimization of the total velocity variation along a boundary (Adelson & Movshon, 1982; Hildreth, 1983). In SBF, however, no boundary is initially given. To solve this problem, we suggested a means, illustrated in Fig. 1(b and c), by which recovery of local orientation informa-

tion and application of a velocity constraint might occur concurrently. Briefly, orientation of an edge moving in some direction δ can be determined from multiple strips of elements that are sequentially affected by the edge. The minimum velocity will occur in the strip perpendicular to the edge [see Fig. 1(c)]. A set of such velocity curves from regions of the moving surface with different orientations can be used to determine the global motion of the surface. The true direction of motion is defined by the intersection of these curves [Fig. 1(c)]. Motion based models differ from the first class of models in that sequences of changes rather than simultaneous spatial positions serve as the basic building blocks of both form and global motion perception in SBF.

Previous research offers no clear basis for preferring either the form-precedes-motion or motion-precedes-form class of models. In Experiment 1, we report an experiment designed to distinguish which class of models operates in SBF. The results clearly support motion-precedes-form models. Based on these results, we develop the general approach offered by Shipley and Kellman (1994) and propose a specific computational scheme for defining local edge segments from motion signals. The proposed account, where local edge orientation is defined by two motion signals, should be particularly sensitive to noise when the two signals are similar in direction and magnitude. Experiment 2 tests this particular motion-precedes-form account by testing for a breakdown in SBF when the disappearances and reappearances are located so that the local motion signals are similar in direction and magnitude.

EXPERIMENT 1

If the initiating conditions for SBF are local motion signals, then adding motion signals that are not triggered by the pseudosurface's motion should disrupt perception of the form of the pseudosurface. In this experiment, we used dots that moved continuously around the array. Such spurious signals should provide local motion signals that could disrupt recovery of other local motion signals and would certainly provide inaccurate inputs for the intersection-of-constraints computation that determines global motion.

In contrast, added moving elements might be expected to have little effect on a form-precedes-motion process. This form defining process relates nearby element changes (STDs) occurring closely in time. In SBF displays, there are no continuously moving points, only spatially and temporally separated element changes, for example, luminance changes. Because the continuously visible moving elements differ from the normal element changes that occur in SBF, the former should not be integrated together with the latter into a form boundary. One might expect a disruptive effect on a form-precedes-motion process if added noise elements do not move in a continuous path, for two reasons. First, such random elements would create element appearances and disappearances that would function as STDs and become integrated with the boundary defining process. Second, a

set of continuously moving points may have a tendency to segregate from the surround, but randomly appearing points might not. Shipley and Kellman (1993a) found that adding STDs at random locations in an SBF display reduced form perception accuracy in SBF; however, when the extra STDs themselves formed a coherent form, the disruptive effect was lessened. Thus, the form-precedes-motion class of models predicts that adding continuously moving elements should have little effect on SBF, whereas adding randomly moving elements should disrupt SBF. A motion-precedes-form model predicts substantial disruption of SBF from spurious motion signals whether these do or do not arise from continuous motion of a coherent form. Experiment 1 investigated the effect of adding a fixed number of spurious motion signals on perception of dynamically defined forms. Forms representing a range of SBF clarity were employed to assure detection of any effect of spurious motion signals. In addition to displays where the spurious signals arose from a stable form, a condition in which random signals were added was also included.

Motion signals were added in three different ways. In one condition, we added these randomly. In two others, we added a group that moved around the screen either in the same or opposite direction to the pseudosurface. Figure 2(a) shows the basic SBF display (the *No Motion* condition) where the pseudosurface translated along a circular path around the center of the screen. In the *Same direction* motion displays, additional elements rotated around the center of the screen in the same direction as the shape to be identified [Fig. 2(b)]; in the *Opposite direction* motion displays, the additional elements rotated in the opposite direction [Fig. 2(c)]; and, in the *Random* motion direction displays, the elements appeared in the same spatial locations as they did in Same and Opposite displays, but in a temporally randomized sequence [Fig. 2(d)]. The Random condition is effectively a random motion condition since the appearance and disappearance of the elements produced apparent motion in all directions.

Experiments 1 and 2 employed an objective form perception task [a 10-alternative forced-choice (AFC) matching task] previously used by Shipley and Kellman (1994) to assess changes in SBF as a function of spatial and temporal display variables. In this task, accuracy increases with texture element density. For Experiment 1 three levels of density were employed covering a four-fold change in density. This assured a broad range of accuracies, increasing the chance of detecting any effect of additional motion signals.

Method

Subjects. Eleven University of Georgia undergraduates served as subjects in 30 min individual testing sessions. Subjects were introductory psychology students who participated for partial fulfillment of course requirements.

Apparatus. All displays were designed and presented using a Macintosh Quadra 800 computer with an E-Machine's TX16, 25 cm high by 33 cm wide RGB

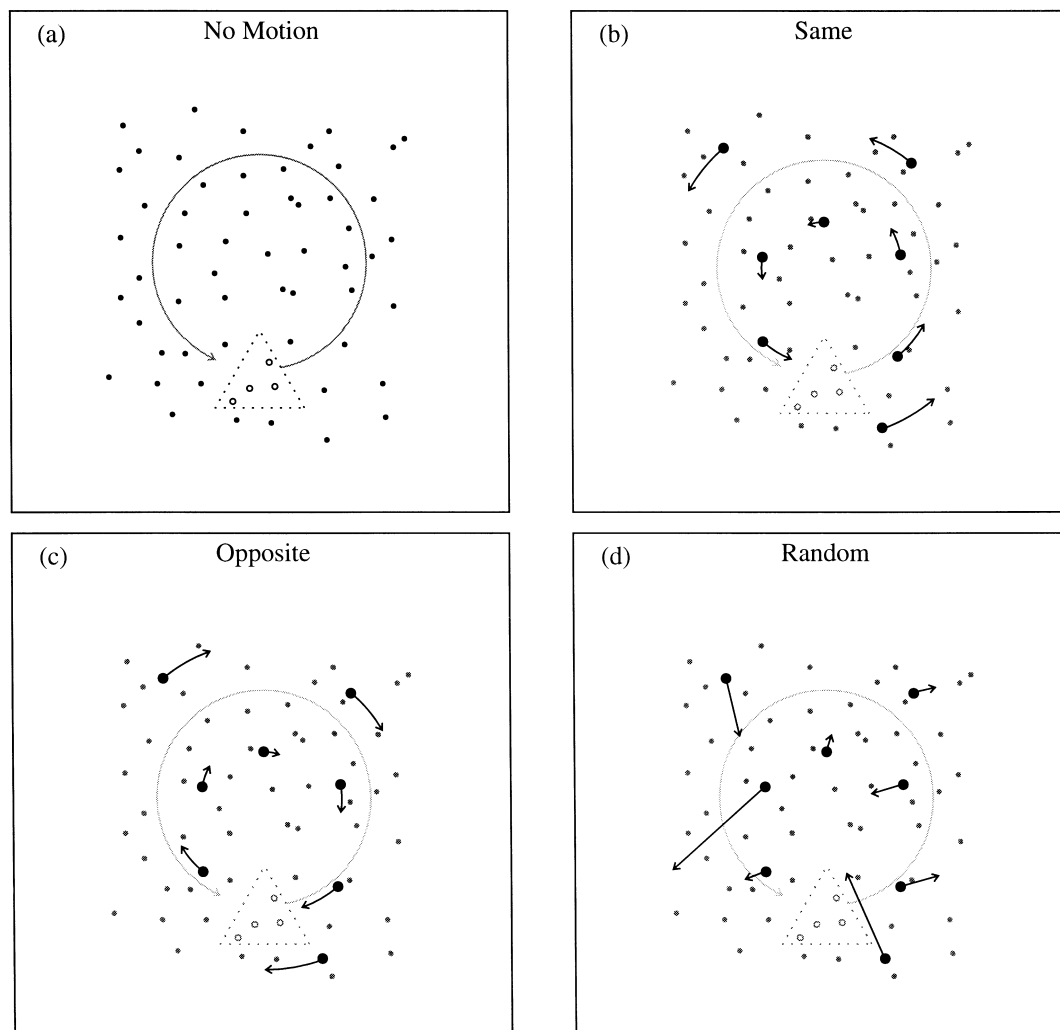


FIGURE 2. (a) An illustration of the pseudosurface (dotted triangle) moving over an element array. In Experiment 1, additional elements were added to this display that all moved in the same direction as the pseudosurface (b), the opposite direction (c), or in random directions (d).

monitor. The screen resolution was 34.25 dots per cm (808 vertical by 1024 horizontal pixels).

Subjects were positioned 150 cm from the monitor. The room was dark except for the illumination provided by the monitor, and a small shielded light (4 W) that illuminated the keyboard so that subjects could enter their responses.

Stimuli. Each display consisted of one of the ten pseudosurfaces shown in Fig. 3. This set was constructed by selecting shapes that did not differ substantially in size, yet provided a range of complexity and discriminability (Shipley & Kellman, 1993c, 1994).

The pseudosurfaces moved over an array of small circles [dia = 0.12 cm (2.68 min arc visual angle)]. Texture density was varied by varying the number of elements placed within a 14.6 × 14.6 cm field (5.58 deg arc visual angle). The number of elements used were 50, 100, and 200. (The display area occupied by elements ranged from 0.2 to 0.8%.) Elements were distributed pseudorandomly within the 14.6 cm square field. To avoid large areas without elements, the distribution was constrained by dividing the field into 100 equal sized

subregions and placing an equal number of elements at random locations within each subregion. For the 50 element condition, 49 subregions were used.

All displays employed unidirectional transformations. As the pseudosurface passed over elements they changed from white (94.6 cd/m²) to black and then returned to white on a black (0 cd/m²) background. This change occurred in a single frame. An element was defined as inside the pseudosurface if the center of the element fell within the pseudosurface, when this occurred the entire element was transformed.

The pseudosurface traveled a circular path with a radius of 3.65 cm (1.39 deg arc). Pseudosurface orientation did not change as it moved over the array. A circular path guaranteed that element transformations would occur with equal frequency along the entire boundary of the pseudosurface and it allowed continuous presentation of the displays.

The *No Motion* displays were generated by selecting 60 equally spaced locations along a circular path [the distance between each location was 0.38 cm (8.7 min arc visual angle)]. These served to position the pseudosurface

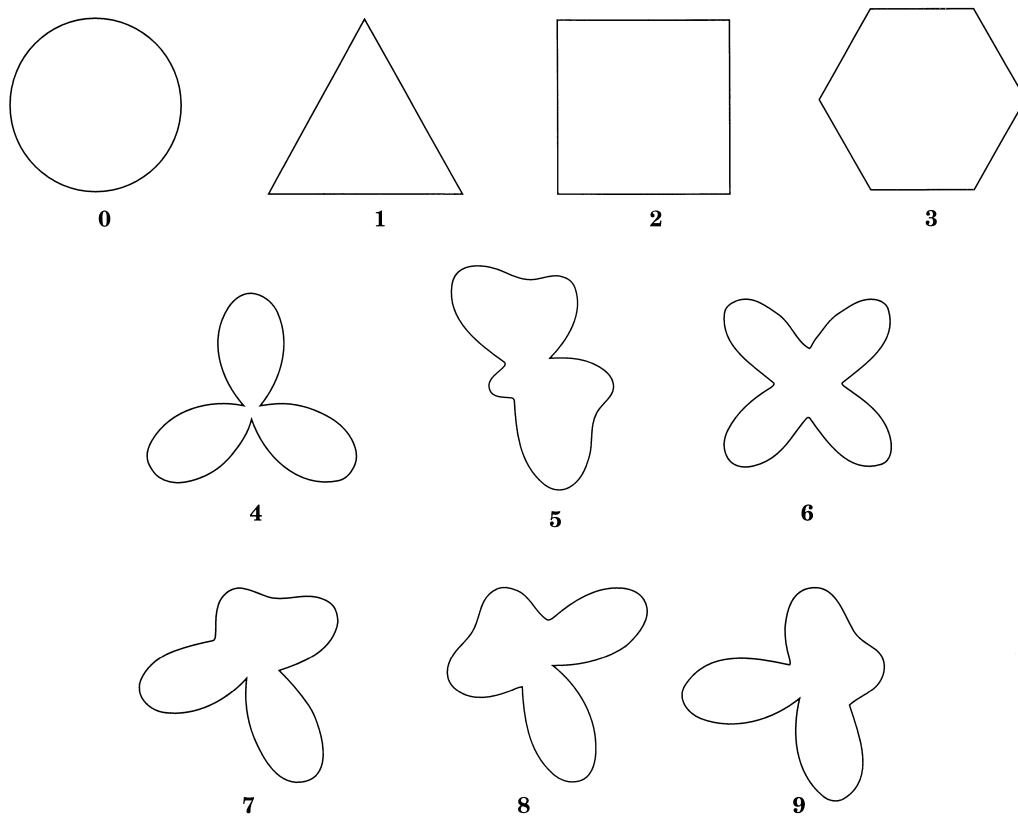


FIGURE 3. The ten forms, or pseudosurfaces, used for the shape identification task in Experiments 1 and 2. [This figure is similar to Fig. 3 from Shipley & Kellman (1994). Copyright © (1994) by the American Psychological Association. Adapted with permission.]

on each of the 60 frames used to animate a display. At each location the elements that were inside the figure were set to black, any that had just left the figure were returned to white. The average number of changes per frame (both white to black and black to white) for 50, 100, and 200 elements was 0.64, 1.35, 2.64.

The *Same* direction displays were constructed by adding eight moving elements to the No Motion displays. Given the size of the pseudosurfaces, eight elements was the minimum number needed to assure that one element would always be near the pseudosurface boundary. In the *Same* condition, these elements circled the array in the same direction, with the same angular velocity, as the pseudosurface. The *Opposite* direction displays used the same element locations employed in the *Same* direction displays but the sequence in which they were shown was reversed. The *Random* direction displays also contained the same eight elements in the same spatial locations as the *Same* and *Opposite* displays but the frame sequence used to animate the moving elements was randomized.

Each display was animated by showing the 60 frames in sequence with each of the frames lasting 33 msec. No interframe interval was used, so the SOA and frame duration were the same. The 60 frame cycle was 2 sec long. Because circular paths were used for both pseudosurfaces and the moving elements, the displays could be presented continuously until the subject responded. When the subject indicated they were ready to identify

the figure, or after 20 complete cycles, the display stopped and subjects entered their selection.

Crossing four motion conditions (None, *Same* direction, *Opposite* direction, and *Random* direction), three element densities, and 10 pseudosurfaces, resulted in 120 displays.

Procedure. The subject's task was a 10-AFC, with accuracy and speed as dependent measures. Subjects had been instructed to indicate their response "as quickly and accurately as possible", and then the 120 displays were presented in random order.

Results

The results of Experiment 1 were clear. Introducing moving elements into an SBF display severely degrades subjects' ability to identify the boundaries of the moving figure. Although both reaction time and accuracy were measured, here we present only accuracy data. Faster reaction times were highly correlated with accuracy. The correlation between mean reaction time and mean accuracy was -0.992 ($P < 0.001$). Overall results are shown in Fig. 4 where mean accuracy of the no motion and the three motion conditions is plotted as a function of texture element density. Although subjects did perform well above chance (10%) in all conditions [all $t(10)s \geq 2.71$, $Ps < 0.03$], accuracies in the three motion conditions were considerably lower than in the No Motion condition. There also appeared to be a small but consistent effect of direction of motion such that performance

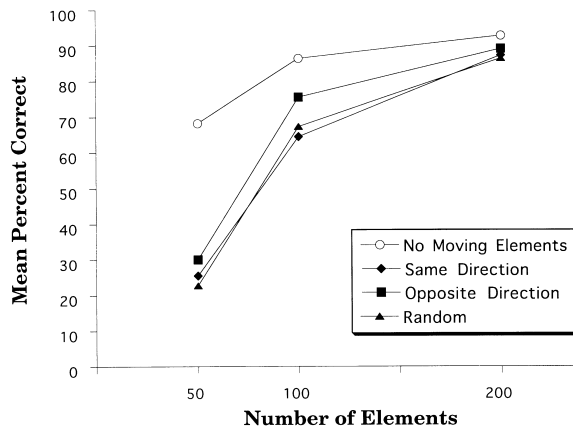


FIGURE 4. Shape identification accuracy for the four conditions in Experiment 1 are plotted as a function of element density.

with Opposite displays was slightly superior to Same and Random displays.

These patterns were confirmed by a two-way ANOVA with motion (No Motion, Same, Opposite, and Random) and number of elements (50, 100, and 200) as within subject factors. Accuracy differed across displays with different motions, $F(3,30) = 37.83$, $P < 0.0001$, and accuracy increased with number of display elements, $F(2,20) = 200.01$, $P < 0.0001$. There was also a significant two way interaction between motion and element density, $F(6,60) = 6.51$, $P < 0.0001$. This interaction may reflect a ceiling effect. In the No Motion condition, accuracies for 100 and 200 elements (86.3 and 92.7%, respectively) did not differ [$F(1,60) = 1.41$, $P > 0.15$], while the other three conditions, which had lower accuracy levels, showed significant increases in accuracy with increases in density [all $F_s(1,60) \geq 6.50$, $P_s < 0.02$].

Accuracy was much higher in the No Motion displays than in the other three. Mean accuracy for the No Motion condition was 82.4%, whereas accuracies for the Same, Opposite, and Random conditions were 59.0, 64.8, and 58.7% respectively [pairwise comparisons between No Motion and the other conditions were significant, all $F(1,30) \geq 47.31$, $P < 0.0001$]. Within the displays with moving elements, Opposite direction motion displays were slightly better than Random and Same [both $F(1,30) \geq 5.07$, $P < 0.05$], while Random and Same did not differ ($F < 1$). Pairwise comparisons at each level of density found that the No Motion condition was superior to the other motion conditions for 50 and 100 elements [$F(1,60) = 31.03$ and 4.16 respectively, both $P < 0.05$], but no differences were found for 200 elements [$F(1,60) = 1.41$, $P > 0.20$]. While the Opposite direction motion condition was consistently slightly more accurate than the Same and Random conditions across density, only for 100 element displays did this difference reach significance [$F(1,60) = 4.16$, $P < 0.05$].

In general, the No Motion condition was superior to the other three for all of the shapes tested. Only at the highest densities, where almost all subjects accurately identified

all shapes, were the four motion conditions comparable. Some shapes, however, appeared to be more stable than others at low element densities. The simple geometric shapes (items 0, 1, and 2 in Fig. 3), were identified more accurately than most of the other forms when the extra motion signals were added. To compare vulnerability to noise, we selected items that were approximately equivalent in recognition accuracy in the absence of noise. On the basis of recognition in the No Motion condition items 4, 5, and 7 were selected (mean accuracies were 83% for items 0–2, and 90% for items 4, 5, 7). With the addition of extra motion signals, accuracy for items 0–2 dropped to 47.8%, while items 4, 5, and 7 dropped to 24.4% [$t(8) = 4.41$, $P < 0.003$].

Identification accuracy reflected the phenomenology of these displays. When accuracy was low no figure was seen, or was seen rarely. Interestingly, when a form was seen in the Same and Opposite conditions, the figure would frequently appear to rotate as it circled the screen. This rotation was perceptually anomalous in that the figure would appear to twist yet not change orientation.

Discussion

Addition of a small number of moving elements had a substantial effect on the SBF process. The result provides the first clear experimental support for motion-precedes-form models. For this class of models, addition of continuously or random moving points should add inappropriate motion signals that disrupt edge recovery. The disruption observed in the Same, Opposite, and Random displays is consistent with this prediction.

For form-precedes-motion models, the addition of continuously moving points should not have had much of an effect. Because of their continuity, individual moving elements would not be predicted to be integrated with other element changes to define boundary segments. In contrast, the elements in the Random condition, which appeared and disappeared at random locations should have disrupted performance. The results of Experiment 1 indicated equivalent performance in the Same and Random conditions, inconsistent with this prediction of form-precedes-motion models. Also, in the Same condition the moving elements defined a rigid moving form. On a form-precedes-motion model, this property would have been expected to reduce intrusions into the pseudosurface form relative to the Random displays where there was no coherent figure available (Shipley & Kellman, 1993a).

The most likely cause of the interference of added motion signals on a motion-precedes-form process is in estimation of a global motion consistent with the various local signals generated around the pseudosurface. Each local motion signaled by successive STDs suffers from an aperture problem (Adelson & Movshon, 1982; Hildreth, 1983) that can be solved utilizing the constraint that the whole pseudosurface has the same motion (Shipley & Kellman, 1994). Spurious motion signals would lead to inaccuracy or indeterminacy in computing global motion, resulting in inaccurate edge integration. The phenomenal

twisting observed in some displays is consistent with distortions in extracting global motion. It is also possible that spurious moving elements interfere by disrupting the extraction of local motion signals as well.

The effect of additional motion signals decreased with increases in element density. Although this may reflect a ceiling effect, it is probable that the decrease also reflected an increase in the signal to noise ratio as texture density increased. Because the number of moving elements was fixed, the signal to noise ratio varied directly with element density. Given a fixed number of noise elements, the proportion of spurious to legitimate motion signals rose as density decreased.

Systematic differences were found between pseudosurface forms in the effect of additional motion signals. Simple geometric forms, such as forms 0–2 in Fig. 3, were more resistant to decrements in performance at lower element densities. This may reflect the smaller amount of information needed to specify these forms. Specifically, it may reflect the interpolation processes that connect locally defined edge pieces in SBF. Connecting local edge segments in SBF might be easiest with smooth edges (*cf.* Kellman & Shipley, 1991). The simple geometric forms (items 0–2) all contain long, smooth edges. Andersen and Cortese (1989) report a similar observation where shape identification in dynamic occlusion displays improved as the number of orientation changes in an object's boundaries decreased.

It is also possible that the familiarity of the simple forms was a factor in boundary stability. Recent work by Shiffrar *et al.* (1997) on the perceptual organization of human stick figures seen walking behind multiple apertures suggests that familiarity may play some role in dynamic unit formation.

EXPERIMENT 2

In Experiment 1 we found that the presence of several motion signals not produced by the appearance and disappearance of texture elements disrupts SBF. Such signals presumably disrupt the integration of motion signals that normally allows concurrent perception of edges and global motion. Shipley and Kellman (1994) showed how edge orientations and global motion can in principle be recovered from several local motion signals, each generated by strips of successive element changes. Here we refine this approach, presenting a particular computational scheme that extracts local boundary orientation from only three successive element changes. This scheme uses the theoretical minimum amount of information required to specify local boundary orientation in the absence of prior information about the figure's global motion.

We label the specific procedure *edge orientation from motion* (EOFM). In this model, the local orientation of an edge is defined by two motion vectors generated by the sequential change of three elements. Figure 5(a) illustrates a series of images where an edge sequentially intersects (either covers or reveals) three elements (labeled 1, 2 and 3 for the order in which they change). Figure 5(b) illustrates the resulting local motion signals. \mathbf{V}_{12} is the vector defined by the spatial and temporal separation of changes in elements 1 and 2, and \mathbf{V}_{23} is the corresponding vector for elements 2 and 3. When \mathbf{V}_{12} and \mathbf{V}_{23} are combined so that they have a common origin, their tips define the orientation of the edge that caused the changes at 1, 2, and 3. (A proof is presented in the Appendix) The true edge motion can then be recovered from the intersection of potential velocities for two such edge segments as described by Shipley and Kellman (1994).

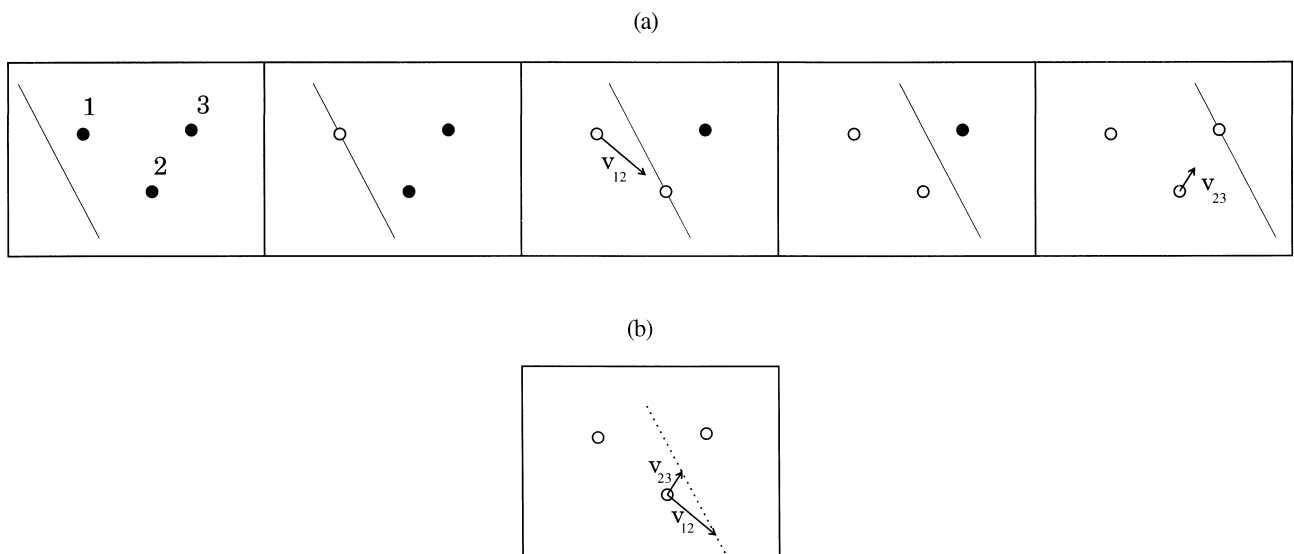


FIGURE 5. An illustration of the local changes that occur as an edge progressively hides or reveals texture elements. In (a), a five frame sequence illustrates an edge sequentially covering three elements. The local motion vectors, \mathbf{V}_{12} and \mathbf{V}_{23} , that are defined by the sequence of disappearances or appearances (illustrated with arrows) can be combined, as shown in (b), to define the orientation of the moving boundary (see text and the Appendix for details).

The final percept, a fully bounded surface, may result from a completion process that connects local edge segments. The edge relationships embodied in the concept of *reliability* (Kellman & Shipley, 1991) may govern completion in this case, as in static edge completion (Shipley & Kellman, 1994). Subjects frequently report that pseudosurfaces which have corners >90 deg appear rounded—subjects confuse a hexagon for a circle but not vice versa. The interpolation process appears to fill in smooth boundaries in such cases. In some displays, however, corners are seen. For example, triangular pseudosurfaces appear to have three corners. Some new concept may be needed to account for completions that contain tangent discontinuities.

EOFM is consistent with the results of Experiment 1 in that combining noise motion vectors with the motion signals defined by element changes would yield incorrect boundary orientation results. By this account, the spatial proximity of the noise motion vectors led to their being combined with the “signal” vectors with the result being unstable local edges.

The approach described here is similar in some respects to one developed by Bruno and Gerbino (1991) for motion defined illusory contours in displays with thin lines. SBF displays do not include any information for boundary position or orientation that might be given by line ends. This allows analysis of the role of spatio-temporal information alone in defining boundaries. Real scenes, however, would certainly

include both element changes and edge changes as observers move. In Bruno and Gerbino's displays the motion signals were coherent and continuous, but it is not clear that their analysis requires either. EOFM requires neither that the signals be coherent nor continuous.

Experiment 1 provided evidence supporting the motion-precedes-form class of models of SBF, and EOFM is a member of this class. But is it the best account available? Can it predict conditions, without extraneous signals, where boundary formation will be weak or absent? While generally robust, the EOFM model is sensitive to noise when the direction and magnitude of the velocity signals are similar. The effect of errors in velocity magnitude estimation on edge orientation is not constant. The size of the effect depends on the relative direction of the two velocity signals: as two signals of similar magnitude approach collinearity, the orientation error increases. Figure 6 illustrates this by plotting the effect of velocity magnitude errors on edge orientation for an edge defined by two equal magnitude velocities. The maximum error in edge orientation (z -axis) is plotted as a function of the relative orientation of the two velocity signals (x -axis), and the size of the velocity error (y -axis). As an example, a 5% error in velocity magnitude (a 5% underestimation of one velocity and a 5% overestimation of the other velocity) results in a 1.6% error in edge orientation when the two equal magnitude vectors are 90 deg apart. In a world where elements are distributed randomly, sequential signals are

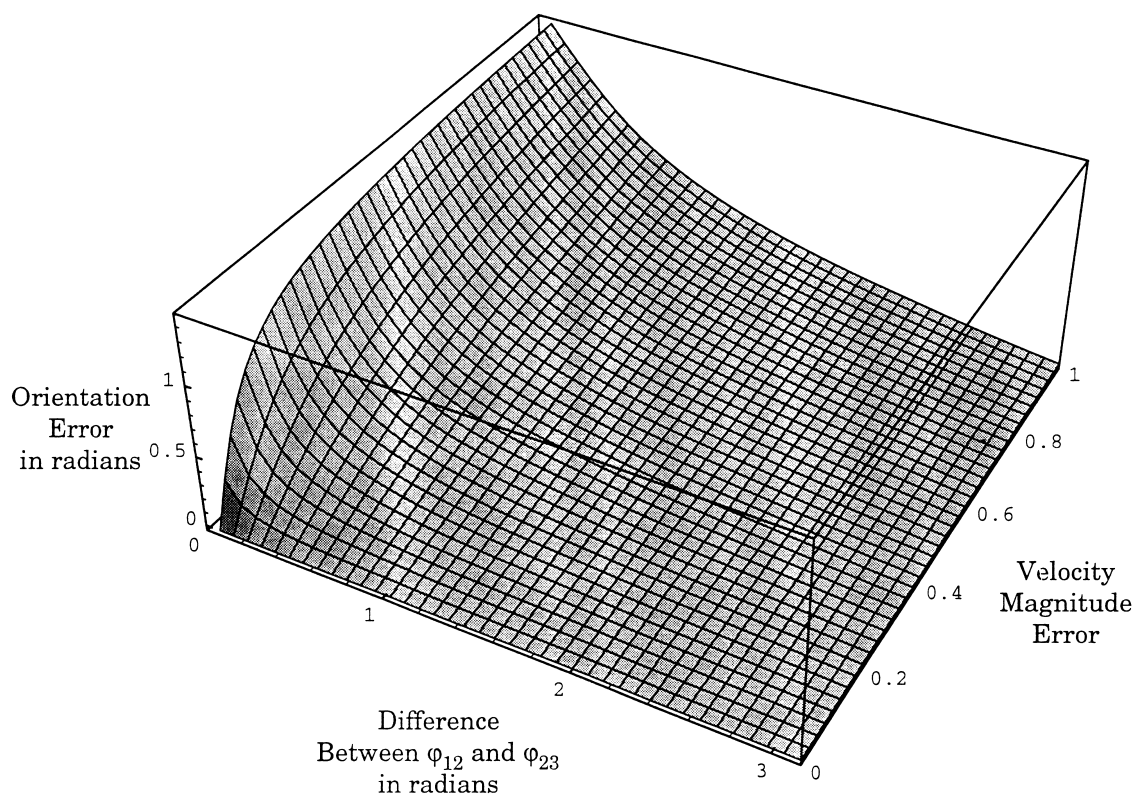


FIGURE 6. Error in edge orientation (z -axis) is plotted as a function of the relative orientation of the two vectors (x -axis) and the magnitude of the velocity error (y -axis) given as a proportion of the signal (a value of 1 indicates that the error is the same magnitude as the signal).

unlikely to be similar, and a mechanism based on local vectors would be quite robust. However, if the sequential signals are similar, boundary formation should be unstable, as small errors in velocity magnitude will result in fluctuations in the perceived form. For example, when the two vectors are 6 deg apart, a 5% velocity error produces a 24.4% (i.e. 43 deg) edge orientation error.

To test the effect of similarity in the direction of the local velocity signals on SBF, displays were created where the changes that occurred along the moving boundary resulted in local motion signals that were similar in magnitude and differed by 6 deg. In these, the *Sequential* displays, the temporally proximal changes were spatially proximal and diverged from collinearity by 6 deg. This was achieved by arranging elements so that each change occurred in sequence at equally spaced locations around the boundary of the form as it moved. In the control, the *Random* displays, changes occurred at the same locations along the boundary of the moving figure, but the sequence of changes was randomized so the local motion vectors were not similar in direction.

Method

Subjects. Ten University of Georgia undergraduates served as subjects in 30 min individual testing sessions. Subjects were introductory psychology students who participated for partial fulfillment of their course requirements.

Apparatus and procedure. The apparatus and procedure were identical to ones used in Experiment 1.

Stimuli. Sequential displays required a new method of generation. Normally, SBF displays are designed by randomly distributing an array of elements, moving a pseudosurface over them, and computing which elements are inside and outside the boundary region on each frame. The displays are then animated by successively changing some property (e.g. color) of elements when they are first encompassed by the pseudosurface and returning that element to its original property when no longer encompassed. It is not possible to precisely constrain

both the spatial and temporal location of element changes using this method because any given element may cross a pseudosurface boundary many times. For the purposes of Experiment 2, texture elements were positioned by sequentially moving the pseudosurface and placing elements at precise points along the pseudosurface's boundaries. The pseudosurface shapes, the path taken by the pseudosurface, the texture elements, and color changes were all identical to the ones employed in Experiment 1.

Figure 7(a) illustrates the entire path of a pseudosurface (the dashed triangle) and the location of the element changes (the small black circles) over 18 frames for a Sequential display. Note that the location of change systematically moves sequentially around the boundary of the pseudosurface. In the actual displays 60 frames (with one or more element changes per frame) were employed, and the sequence of element changes was divided into six subsequences (the reasons for this are elaborated below).

This procedure generates displays in which element changes were located so that sequential velocity signals have similar magnitude and differ by 6 deg. One property of these displays is that each element changes only once during each cycle, however, a given element may cross the boundary of the moving figure several times. Although the appearance of a large number of stationary elements inside the moving form degrades performance, particularly at low densities (Cunningham *et al.*, 1996), this property was identical for both sequential and random displays. The control (Random) displays were generated using the same locations within the pseudosurface employed for the Sequential displays, but the sequence of locations was randomized across frames [Fig. 7(b)].

In Fig. 7(a) the black elements may be perceptually grouped into a continuous line. When more frames are employed, the figures defined by these lines resemble the form of the pseudosurfaces that creates them. We were concerned that subjects might base their responses on

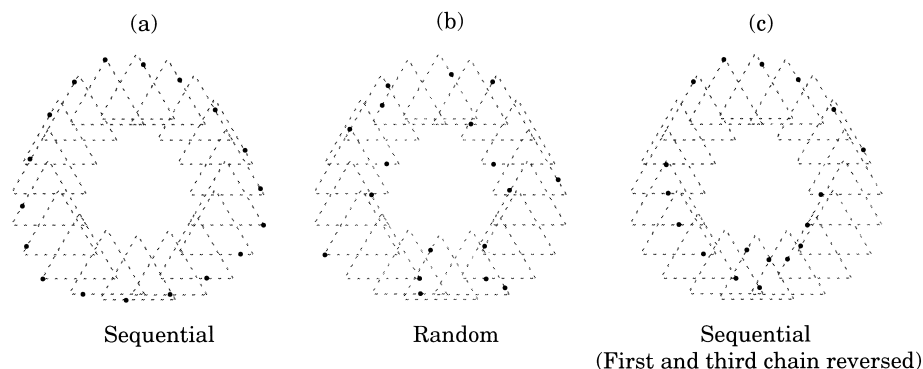


FIGURE 7. Illustrations of the changes in the two types of SBF displays used in Experiment 2. In each illustration the pseudosurface (dashed triangle) is shown at each of the locations that it passes through as it moves on a circular path. In (a), the *Sequential* displays, each change (illustrated with black dots) occurred near the last change. In (b), the *Random* displays, each change occurred at the same locations within the triangle but at random positions within the sequence. In (c) the chains of element changes are reorganized so the final third of the triangle is specified first, the second third second, and the first third last; see text for details on display construction.

these patterns, rather than trying to report the moving figure. Such an artifact would reduce, not enhance the hypothesized effect in the sequential condition. To eliminate this possibility, the entire sequence of element changes was broken up into six chains of ten sequential element changes. By varying the order of the chains [Fig. 7(c)] it was possible to present changes that produced locally similar velocity signals (within each chain), and defined the entire boundary across chains, but did not provide any extraneous cue to the pseudosurface's shape. This manipulation did introduce some nonsimilar motion signals into the Sequential displays at the point where each chain ended. These few motion signals might have weakened the experimental manipulation slightly; however, most motion signals in the Sequential condition were nearly collinear and were predicted to cause ample disruption of the SBF process based on the EOFM computational scheme.

To generate displays which ranged in difficulty, the number of element changes per frame was varied. In each display, one, two, or four elements changed per frame. When more than one element changed in a frame the locations of the changes were positioned to be maximally distant along the boundary of the pseudosurface.

This procedure for generating SBF displays only determines the location of texture elements within a roughly doughnut shaped region defined by the spatio-temporal overlap of each pseudosurface with the background (see Fig. 7). To create a square element field like the one used in Experiment 1, the 5.58 deg region was divided into 100 equal sized subregions and additional elements were added in random location. These elements were added in such a way that each subregions in the one, two, and four changes per frame displays would have at least one, two, or four elements, respectively.

Finally, both unidirectional and bidirectional changes were employed. In unidirectional sequential displays, elements within a chain would sequentially appear or disappear depending on whether the elements were at leading or trailing edges. In contrast, in unidirectional random displays, disappearance and reappearance could alternate from one frame to the next. In order to include sequential displays with alternating appearance and disappearance, we also included displays with bidirectional element changes. In bidirectional displays, the color of the element and the location along the moving edge are not correlated (e.g. half of the changes along the trailing edge would be from black to white, and half from white to black). In the bidirectional sequential displays appearance and disappearance alternated within each chain of elements.

The animation procedure was the same as that for Experiment 1, except that before each 60-frame sequence the screen was briefly cleared.

Crossing Random and Sequential changes, Unidirectional and Bidirectional changes, one, two, or four changes per frame, and 10 pseudosurfaces, resulted in 120 displays.

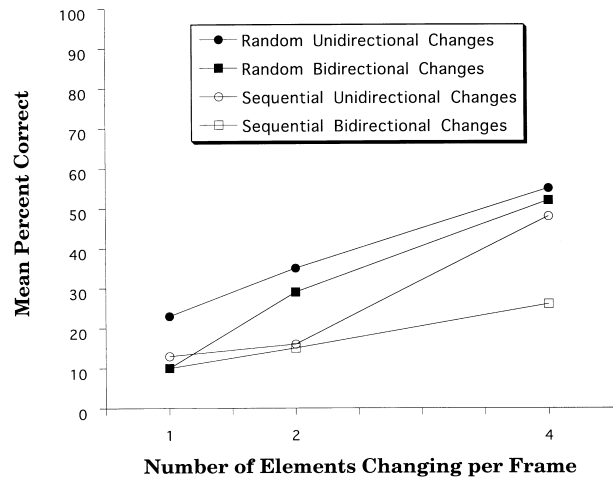


FIGURE 8. Shape identification accuracy for the four conditions in Experiment 2 are plotted as a function of number of element changes per frame.

Results

The results of Experiment 2 are shown in Fig. 8 where accuracy is plotted as a function of number of element changes per frame. Subjects often identified the shapes in the Random condition (overall accuracy = 43.8%) but were much less accurate with Sequential displays (overall accuracy = 23%).

The phenomenal appearance of the two types of displays was noteworthy. Most of the Sequential displays did not appear to contain a moving figure. They did appear to contain one or more moving elements. In contrast, a moving figure was always apparent in the Random displays.

The superiority of the Random condition was confirmed by a three-way ANOVA with Sequence (Random vs Sequential), direction of element change (Unidirectional vs Bidirectional), and number of element changes per frame as within subject factors. Accuracy was markedly higher in Random displays than Sequential displays, $F(1,9) = 26.40$, $P < 0.001$, Unidirectional displays were better than Bidirectional displays, $F(1,9) = 32.00$, $P < 0.001$, and accuracy increased with number of element changes per frame, $F(2,18) = 72.29$, $P < 0.0001$.

The interaction term between number of element changes and Sequence type was significant, $F(2,18) = 8.98$, $P < 0.01$. Furthermore, the interaction term between number of element changes and direction of element changes was marginally significant, $F(2,18) = 2.68$, $P < 0.1$. In both cases, accuracy for one change per frame were the lowest terms (12% for one-change Bidirectional displays and 12.5% for one-change Sequential displays) and did not significantly differ from chance (both $t(19) \leq 1.31$, $P > 0.20$), so both interaction terms probably reflect a floor effect. The other two interaction terms (Sequence type by direction of element change and the three-way term) were not significant, both $F < 1$.

The effect of sequence type was generally seen across all displays. Employing pairwise comparisons between Random and Sequential displays, all but one random display was found to be significantly higher than the corresponding Sequential display [all $F(1,18) \geq 6.45$, $P < 0.02$]. The one-change per frame bidirectional displays did not differ, however, these accuracies were also not significantly greater than chance [both $t(9) < 1.16$, $P > 0.20$].

Discussion

Similarity in the direction of sequential element changes clearly disrupted SBF. This finding is consistent with the computational approach we have outlined. Edges defined by sequential motion signals of similar direction and magnitude would be inherently unstable, sensitive to any error in the detected magnitude of the signals. When positioned randomly, so that sequential signals were not similar in direction, stable edges were seen.

The motion-precedes-form model of SBF presented earlier may also account for the phenomenal appearance of the Sequential displays. Here subjects frequently saw moving elements, not a global form. This may reflect an important relationship between perception of boundaries and perception of motion, which we consider below.

GENERAL DISCUSSION

In SBF, continuous boundaries, form and global motion are perceived from local element changes sparsely distributed in space and time. SBF thus represents an interesting feat of perceptual organization, but more importantly, it reveals computational strategies employed by the visual system for determining object boundaries and spatial layout, especially in the face of fragmentary information (Shepard, 1984; Shipley & Kellman, 1994). The use of motion signals in the determination of occlusion boundaries may also reflect the premium placed on detection of motion in the world. Since retinal motion is not necessarily linked to motion of objects, the visual system needs a way to determine whether the motion energy present at the retina is a result of an object moving or an edge occluding a surface. The conditions that do and do not trigger SBF reflect the use of sets of local motion signals to define surface boundaries under certain conditions and to signal actual element motions under others.

Motion-precedes-form models of SBF thus avoid potential confusion in determining motion, and provide a robust process for determining surface shapes. The present experiments provide direct evidence for the motion-precedes-form class of models. The initial stage of processing is extraction of local motion signals based on pairs of element changes (STDs). Then global motion and boundary orientation are derived from these signals. Experiment 1 tested the general motion-precedes-form idea by adding spurious local motion signals. Displays were structured so that a form-precedes-motion process should not have been greatly perturbed by this manipula-

tion. Disruption of shape identification performance supports motion-precedes-form models.

Experiment 2 tested a specific motion-precedes-form computation where boundary orientation in a local region is obtained from two motion vectors. This scheme, while generally robust, is sensitive to errors when successive STDs are nearly collinear. The results of Experiment 2 provided support for this computation: shape identification was poor in displays with sequential motion signals with similar orientations, and much better in displays where the signal directions were distributed randomly.

The motion-precedes-form model of SBF provides an account of the performance data, and may also make sense of the phenomenal appearance of the Sequential and Random displays in Experiment 2. A fascinating aspect of displays that produce SBF is that they give rise to little or no apparent motion of elements. Successive STDs in nearby elements would be predicted to produce apparent motion between their locations according to models of the correspondence process for apparent motion (e.g. Ullman, 1979). When all of the local STDs are consistent with a single moving object, however, the local signals lead to perception of that object and its global motion rather than local motions of or between elements. A number of researchers have noted an inverse relationship between seeing a boundary and seeing individual elements move (Bruno & Gerbino, 1991; Petersik & McDill, 1981; Shipley & Kellman, 1993b, 1994). Perception of boundaries and perception of local motion are complementary. Integrating local motion signals to define a moving boundary prevents the local signals from having a phenomenal affect. When not integrated, as in the Sequential displays, they appear as element motion.

It is interesting to consider why this visual mechanism might exist, given that in natural scenes, object segmentation and motion perception are normally supported by additional information, such as luminance, texture, and depth differences between visible surfaces. One answer is that such differences are occasionally lacking in the optic array. SBF allows perception of boundaries, form, and motion in the absence of these information sources so long as sequential changes in sparse visible elements are available. Such changes will generally be available whenever there is relative motion between surfaces, whether the nearer surface is opaque or translucent. As such the SBF process is completely general, and may serve as the basis for the perception of stable surface qualities over time.

A major challenge in visual science is to connect the mechanisms responsible for generating local motion signals with computations of various perceptual outcomes that depend on higher-order information. For example, triggering of a velocity-sensitive cell in MT does not imply that any viewed object has moved or has been perceived as moving. Such a cell might respond either when the observer views a moving target or when a moving observer views a stationary target. Another example is the resolution of aperture problems (e.g.

Hildreth, 1983), here local motions belonging to edges of a connected region lead to a determination of a coherent global motion. SBF takes this complexity a step further, because no connected boundaries or edge orientations are directly given in the optic array. From a sequence of small elements appearing and disappearing a stable continuous edge is seen moving over an array of stationary elements.

The computational approach we have outlined fits with some of the known components of early visual processing. We have characterized the various initiating conditions (element changes) for SBF as STDs. All of the element changes that we have observed to produce spatio-temporal boundaries include luminance changes at discrete locations (Shipley & Kellman, 1993c, 1994). These STDs may be picked up by the Y cells in the LGN. These cells are generally considered the input for motion sensitive units in MT. There appear to be two general types of motion sensitive cells in MT (Tanaka *et al.*, 1986). Cells respond to either motion in a restricted spatial region, or motion over a wide region. The former have been termed MT_o (object) cells and the latter MT_f (field) cells. This general division fits with a division in the information provided by motion in the two cases. Shipley and Kellman (1993c) argued for a division between information that specifies observer motion, usually referred to as optic flow, and information for surface boundaries, which we termed *optic tearing*. The rationale for the division was that the optical change information in the two cases differ in kind, despite the fact both are motion based. Observer or object motions typically result in regions of homogeneous local motion signals along continuous paths. In contrast, dynamic occlusion may result in local motion signals with varied position and directionality. While local motion signals may be used in both cases, the motion signals generated at edges could interfere with computations of observer motion, and conversely motion signals from observer motion could interfere with boundary computations. While the MT_f-MT_o division makes functional sense, the use "object" for MT_o cells may be misleading. Activity in such a cell may result in the perception of a single moving object, or as suggested here, if the activity occurs in conjunction with activity in other cells it may contribute to the perception of an occluding edge.

Our improving understanding of the process that extracts boundary, form, and global motion in dynamic displays, and of when the process breaks down, may help lead toward a general account of how the segmentation of scenes, the coherence of objects, and the motions of objects and the observer are computed concurrently from optical change information.

REFERENCES

- Adelson, E. H. & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300, 523–525.
- Andersen, G. J. & Cortese, J. M. (1989). 2-D contour perception resulting from kinetic occlusion. *Perception and Psychophysics*, 46, 49–55.
- Bruno, N. & Bertamini, M. (1990). Identifying contours from occlusion events. *Perception and Psychophysics*, 48, 331–342.
- Bruno, N. & Gerbino, W. (1991). Illusory figures based on local kinematics. *Perception*, 20, 259–274.
- Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, 5, 1986–2007.
- Cunningham, D. W., Shipley, T. F. & Kellman, P. J. (1996). Spatio-temporal boundary formation: The role of global motion signals. *Investigative Ophthalmology and Visual Science*, 37 (Suppl.), S172.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gibson, J. J. (1968). *The change from visible to invisible: A study of optical transitions* [Film]. Pennsylvania: Psychological cinema register, State College.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale: Lawrence Erlbaum.
- Gibson, J. J., Kaplan, G. A., Reynolds, H. N. & Wheeler, K. (1969). The change from visible to invisible: A study of optical transitions. *Perception and Psychophysics*, 5, 113–116.
- Hildreth, E. C. (1983). *The measurement of visual motion*. Cambridge: The MIT Press.
- Kellman, P. J. & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23, 141–221.
- Petersik, J. T. & McDill, M. (1981). A new bistable motion illusion based upon "kinetic optical occlusion". *Perception*, 10, 563–572.
- Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imaging thinking and dreaming. *Psychological Review*, 91, 417–447.
- Shiffrar, M., Lichley, L. & Chatterjee, S. H. (1997). The perception of biological motion across apertures. *Perception and Psychophysics* in press.
- Shipley, T. F. & Kellman, P. J. (1993a). Competition and cooperation in spatiotemporal boundary formation. Paper presented at the 34th Annual Meeting of The Psychonomic Society, Washington, DC.
- Shipley, T. F. & Kellman, P. J. (1993b). Spatiotemporal boundary formation: Temporal integration is confined to a 150 msec window. *Investigative Ophthalmology and Visual Science*, 34 (Suppl.), 1082.
- Shipley, T. F. & Kellman, P. J. (1993c). Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form and global motion? *Spatial Vision*, 7, 323–339.
- Shipley, T. F. & Kellman, P. J. (1994). Spatiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General*, 123, 3–20.
- Stappers, P. J. (1989). Forms can be recognized from dynamic occlusion alone. *Perceptual and Motor Skills*, 68, 243–251.
- Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y. & Iwai, E. (1986). Analysis of the local and wide-field movements in the superior temporal visual areas of the macaque monkey. *Journal of Neuroscience*, 6, 134–144.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge: MIT Press.

Acknowledgements—This research was supported by NSF Research Grants BNS 91-20919 to TFS, and SBR-9496112 to PJK. Portions of this research were presented at the 1994 Annual Meeting of The Association for Research in Vision and Ophthalmology. We thank Robert Futamura and Douglas W. Cunningham for help in creating the displays, collecting the data, and thoughtful criticism. We would also like to thank Ellen Hildreth and two anonymous reviewers for their thoughtful comments on an earlier draft of the paper.

APPENDIX

In Fig. A1, the distances (B_{12} and B_{23}) traveled by the edge as it sequentially affects elements 1, 2, and 3, are a product of the time

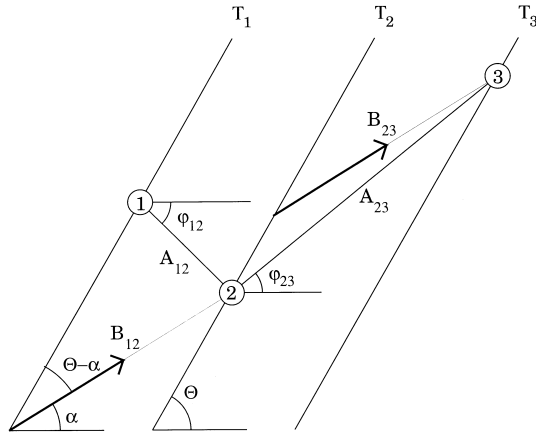


FIGURE A1. An illustration of an edge sequentially covering three elements at times T_1 , T_2 , and T_3 . The circles numbered 1, 2, and 3 represent the three elements. θ is the orientation of the edge. α is the direction of motion of the edge (indicated by the arrows along the lines B_{12} and B_{23}). B_{12} and B_{23} are the distances traveled by the edge between elements 1 and 2 and 2 and 3, respectively. A_{12} and A_{23} are the distances between elements 1 and 2 and 2 and 3, respectively. φ_{12} and φ_{23} are orientations of the lines A_{12} and A_{23} , respectively (these represent the directions of the two local motion signals defined by the sequence of changes at 1, 2, and 3).

between changes and the velocity of the edge (V). In Eqs (A1) and (A2) ΔT_{12} and ΔT_{23} refer to the difference in time between the first and second element change ($T_2 - T_1$), and the second and third element changes ($T_3 - T_2$), respectively.

$$B_{12} = V * \Delta T_{12} \quad (A1)$$

$$B_{23} = V * \Delta T_{23} \quad (A2)$$

In Eqs (A1) and (A2), B_{12} , B_{23} , and V are all unknowns. However, using the Law of Sines [Eqs (A3) and (A5)] the distance traveled by the edge can also be expressed as a function of the distance between points and the orientation and direction of motion of the edge [Eqs (A4) and (A6)].

$$\frac{B_{12}}{\sin(180 - (\theta - \varphi_{12}))} = \frac{B_{12}}{\sin(\theta - \varphi_{12})} = \frac{A_{12}}{\sin(\theta - \alpha)} \quad (A3)$$

$$B_{12} = \frac{A_{12} * \sin(\theta - \varphi_{12})}{\sin(\theta - \alpha)} \quad (A4)$$

$$\frac{B_{23}}{\sin(180 - (\theta - \varphi_{23}))} = \frac{B_{23}}{\sin(\theta - \varphi_{23})} = \frac{A_{23}}{\sin(\theta - \alpha)} \quad (A5)$$

$$B_{23} = \frac{A_{23} * \sin(\theta - \varphi_{23})}{\sin(\theta - \alpha)} \quad (A6)$$

Substituting Eqs (A1) and (A2) into Eqs (A4) and (A6) and solving for V gives Eqs (A7) and (A8).

$$V = \frac{A_{12}}{\Delta T_{12}} * \frac{\sin(\theta - \varphi_{12})}{\sin(\theta - \alpha)} \quad (A7)$$

$$V = \frac{A_{23}}{\Delta T_{23}} * \frac{\sin(\theta - \varphi_{23})}{\sin(\theta - \alpha)} \quad (A8)$$

Assuming the velocity of the moving edge is constant allows Eqs (A7) and (A8) to be combined into Eq. (A9), which can be reduced to Eq. (A10).

$$\frac{A_{12}/\Delta T_{12}}{A_{23}/\Delta T_{23}} = \frac{\sin(\theta - \varphi_{23})}{\sin(\theta - \alpha)} \bigg/ \frac{\sin(\theta - \varphi_{12})}{\sin(\theta - \alpha)} \quad (A9)$$

$$\frac{A_{12}/\Delta T_{12}}{A_{23}/\Delta T_{23}} = \frac{\sin(\theta - \varphi_{23})}{\sin(\theta - \varphi_{12})} \quad (A10)$$

If we define $v_{12} = A_{12}/\Delta T_{12}$ and $v_{23} = A_{23}/\Delta T_{23}$ as the local velocities for motion between the sequentially changing points, and K to be the ratio of these such that $K = v_{12}/v_{23}$, then Eq. (A10) can be rewritten as:

$$\begin{aligned} K * \sin \theta * \cos \varphi_{12} - K * \cos \theta * \sin \varphi_{12} \\ = \sin \theta * \cos \varphi_{23} - \cos \theta * \sin \varphi_{23} \end{aligned} \quad (A11)$$

Rearranging terms:

$$\sin \theta * (\cos \varphi_{23} - K * \cos \varphi_{12}) = \cos \theta * (\sin \varphi_{23} - K * \sin \varphi_{12}) \quad (A12)$$

$$\tan \theta = \frac{\sin \varphi_{23} - K * \sin \varphi_{12}}{\cos \varphi_{23} - K * \cos \varphi_{12}} \quad (A13)$$

Finally, Eq. (A14) is an equation for the orientation of the edge as a function of observable quantities: the relative position of the elements, time between changes, and distances between changing elements:

$$\theta = \tan^{-1} \left(\frac{\sin \varphi_{23} - K * \sin \varphi_{12}}{\cos \varphi_{23} - K * \cos \varphi_{12}} \right) \quad (A14a)$$

or

$$\theta = \tan^{-1} \left(\frac{v_{23} * \sin \varphi_{23} - v_{12} * \sin \varphi_{12}}{v_{23} * \cos \varphi_{23} - v_{12} * \cos \varphi_{12}} \right) \quad (A14b)$$